Journal Pre-proof

Data-Driven Multiscale Modeling in Mechanics

K. Karapiperis, L. Stainier, M. Ortiz, J.E. Andrade

PII: S0022-5096(20)30453-1
DOI: https://doi.org/10.1016/j.jmps.2020.104239
Reference: MPS 104239

To appear in: *Journal of the Mechanics and Physics of Solids*

Received date: 14 July 2020
Revised date: 14 November 2020
Accepted date: 15 November 2020

Please cite this article as: K. Karapiperis, L. Stainier, M. Ortiz, J.E. Andrade, Data-Driven Multiscale Modeling in Mechanics, *Journal of the Mechanics and Physics of Solids* (2020), doi: https://doi.org/10.1016/j.jmps.2020.104239

Data-Driven Multiscale Modeling in Mechanics

K. Karapiperis[a], L. Stainier[b], M. Ortiz[c], J. E. Andrade[a,*]

[a] *Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA*
[b] *Institut de Recherche en Génie Civil et Mécanique (GeM), Ecole Centrale de Nantes, 4321 Nantes cedex 3, France*
[c] *Graduate Aerospace Laboratories, California Institute of Technology, Pasadena, CA 91125, USA*

Abstract

We present a Data-Driven framework for multiscale mechanical analysis of materials. The proposed framework relies on the Data-Driven formulation in mechanics [1], with the material data being directly extracted from lower-scale computations. Particular emphasis is placed on two key elements: the parametrization of material history, and the optimal sampling of the mechanical state space. We demonstrate an application of the framework in the prediction of the behavior of sand, a prototypical complex history-dependent material. In particular, the model is able to predict the material response under complex nonmonotonic loading paths, and compares well against plane strain and triaxial compression shear banding experiments.

Keywords: data-driven mechanics, multiscale modeling, granular materials

1. Introduction

Traditionally, the mechanical behavior of history-dependent materials has been described by empirical constitutive laws formulated within the framework of continuum thermodynamics and plasticity theory [2–4]. In that context, the material state is described by internal variables that are often phenomenological, and are subject to *ad-hoc* evolution laws (kinetic relations). Advancements in atomistic [5, 6] and micromechanical [7, 8] simulation have inspired physics-based internal variables that encapsulate the microstructure (e.g. density of defects/dislocations for heterogeneous solids [9, 10], fabric tensors for granular materials [11], etc). Naturally, these developments also gave rise to multiscale methods, which attempt to either pass information from the fine to the coarse scale (hierarchical) [12–15] or seamlessly connect the two scales by modeling their interaction (concurrent/semi-concurrent) [16–20].

Despite the relative success of these conventional paradigms, further progress has been hindered by many challenges. In the case of conventional constitutive modeling, the process inherently induces uncertainty due to the imperfect knowledge of the functional form of the constitutive laws [1], and their extrapolative properties beyond the finite data set used for calibration. Moreover, the process of calibration itself can be challenging due to the continuously increasing model complexity (e.g. [21, 22]). On the other hand, conventional multiscale methods can be notoriously demanding in computational time and memory. The data obtained from lower-scale simulations, carried out concurrently or on-the-fly, are never reused, rendering the conventional paradigm inherently inefficient.

In light of the aforementioned challenges, and driven by the progress in Data Science, promising alternatives have surfaced in the form of machine learning and Data-Driven techniques. The first machine learning studies relied on neural networks (NN) trained with experimental data in

*Corresponding author
Email address: jandrade@caltech.edu (J. E. Andrade)

order to predict mechanical behavior [23]. With the increase in computation, deep learning has found multiple applications in behavior prediction with tensor basis- [24], hybrid graph- [25] and recurrent-NN [26, 27], used as surrogates for constitutive laws or lower-scale simulations. Despite their efficiency and, in some cases, their desired built-in invariance properties, the above machine learning techniques leave a lot to be desired. They rely on a (hidden) mathematical representation of constitutive relations, therefore leading to extrapolation from the training material data set, and often depend on application-specific network architectures.

The Data-Driven paradigm for computational mechanics [1, 28] bypasses any modeling step, by formulating the problem directly on a given material data set, while enforcing pertinent constraints and conservation laws. More specifically, by defining a phase space of stress-strain field pairs, this approach leads to a distance minimization problem between a given material data set and the subset of field pairs that satisfy the conservation laws. Recently, the Data-Driven paradigm was extended to dynamics [29] and inelasticity [30]. Also, variations of the Data-Driven framework have been proposed that consider locally linear [31] or locally convex embeddings [32].

In this study, we present a new multiscale Data-Driven paradigm as a complement to conventional multiscale modeling. The framework departs from previous developments in Data-Driven computing in that i) material data are directly extracted from lower-scale computations (e.g. molecular dynamics (MD), discrete element method (DEM), finite element method (FEM)), ii) it enables optimal phase space sampling, offline or on-the-fly, and iii) it is thermodynamically consistent without imposing any constraints on the nature of history-dependence.

The paper is organized as follows. In Section 2 we review the general framework of Data-Driven Inelasticity, which represents our study's point of departure. Section 3 addresses phase space sampling from lower-scale computations, and briefly introduces optimal strategies. Next, in Section 4, we discuss pertinent general representations of the material history that ensure the consistency of the multiscale computation. Finally, Section 5 presents a detailed application of the method in multiscale modeling of granular materials. We conclude with a discussion of the proposed framework in Section 6.

2. Data-Driven Inelasticity

Consider the purely mechanical problem of a body that is discretized in N nodes and M material points (Fig 1 a)). The body is subject to applied forces $\mathbf{f} = \{\mathbf{f}_\alpha\}_{\alpha=1}^N$, and undergoes displacements $\mathbf{u} = \{\mathbf{u}_\alpha\}_{\alpha=1}^N$ at its nodes. The internal state is characterized by local stress and strain pairs $\{(\boldsymbol{\varepsilon}_e, \boldsymbol{\sigma}_e)\}_{e=1}^M$. We consider the pair $\mathbf{z}_e = (\boldsymbol{\varepsilon}_e, \boldsymbol{\sigma}_e)$ as a point in a local phase space Z_e, and $\mathbf{z} = \{\mathbf{z}_e\}_{e=1}^M$ as a point in the global phase space Z. Within a time-discrete formulation, the internal state of the system is subject to the following compatibility and equilibrium constraints:

$$\boldsymbol{\varepsilon}_{e,k} = \mathbf{B}_e u_k, \quad e = 1, \dots, M \tag{1}$$

$$\sum_{e=1}^{M} w_e \mathbf{B}_e^T \boldsymbol{\sigma}_{e,k} = \mathbf{f}_k \tag{2}$$

where $\mathbf{u}_k, \mathbf{f}_k, \boldsymbol{\varepsilon}_k, \boldsymbol{\sigma}_k$ denote the displacements, forces, strains and stresses at time t_k respectively, while $\{w_e\}_{e=1}^M$ are elements of volume, and $\mathbf{B}_e$ is a discrete strain operator for material point e. A constraint set E_k may be defined as:

$$E_k = \{\mathbf{z} \in \mathbf{Z} \,|\, (1) \text{ and } (2)\,\} \tag{3}$$

and consists of all the compatible and equilibrated internal states, due to the applied loads and displacements at time t_k.

Instead of providing closure to Equations 1 and 2 by postulating a constitutive relation $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\varepsilon})$, the Data-Driven formulation of the mechanical problem relies directly on the material data. More specifically, the material behavior is described by a material data set $D_{e,k}$ of points $\mathbf{z}_{e,k} \in Z_{e,k}$, that is attainable at a material point given its past local history of deformation:

$$D_{e,k} = \{(\boldsymbol{\varepsilon}_{e,k}, \boldsymbol{\sigma}_{e,k}) \,|\, (\text{past local history}) \} \tag{4}$$

Accordingly, a global material data set may be defined as:

$$D_k = D_{1,k} \times \ldots \times D_{M,k}$$

The history-dependent Data-Driven problem consists of finding the compatible and equilibrated internal state $\mathbf{z}_k \in E_k$ that minimizes the distance to the global material data set D_k at time t_k. Equivalently, the problem consists of finding the point $\mathbf{y}_k$ in the material data at time that is closest to the constraint set E_{k+1} at time t_k. To this end, the local phase spaces Z_e are equipped with the following metric:

$$|\mathbf{z}_e|_e = \left(\mathbb{C}_e \boldsymbol{\varepsilon}_e \cdot \boldsymbol{\varepsilon}_e + \mathbb{C}_e^{-1} \boldsymbol{\sigma}_e \cdot \boldsymbol{\sigma}_e\right)^{1/2} \tag{5}$$

where $\mathbb{C}_e$ are symmetric positive definite matrices, that are only introduced as a numerical scheme, and do not represent actual material behavior. The above norm induces a metrization of the global phase space Z, by means of the norm:

$$|\mathbf{z}| = \Big(\sum_{e=1}^{M} w_e |\mathbf{z}_e|_e^2 \Big)^{1/2} \tag{6}$$

with corresponding global distance:

$$d(\mathbf{z}, \mathbf{y}) = |\mathbf{z} - \mathbf{y}| \tag{7}$$

Consisely, the time discrete Data-Driven problem may be formulated as:

$$\min_{\mathbf{y} \in D_k} \min_{\mathbf{z} \in E_k} d(\mathbf{z}_k, \mathbf{y}_k) = \min_{\mathbf{z} \in E_k} \min_{\mathbf{y} \in D_k} d(\mathbf{z}_k, \mathbf{y}_k) \tag{8}$$

For fixed $\mathbf{y}_k \in D_k$, the closest point projection $\mathbf{z}_k = P_{E_k} \mathbf{y}_k$ involves minimizing the quadratic function $d^2(\cdot, \mathbf{y}_k)$ subject to the compatibility and equilibrium constraints (Eqs 1-2). The former may be enforced directly, while the latter may be enforced by means of Lagrange multipliers $\boldsymbol{\eta}_k$, which represent virtual displacements of the system. The resulting Euler-Lagrange equations read:

$$\Big(\sum_{e=1}^{M} w_e \mathbf{B}_e^T \mathbb{C}_e \mathbf{B}_e \Big) \mathbf{u}_k = \sum_{e=1}^{M} w_e \mathbf{B}_e^T \mathbb{C}_e \boldsymbol{\varepsilon}_{e,k}^* \tag{9}$$

$$\Big(\sum_{e=1}^{M} w_e \mathbf{B}_e^T \mathbb{C}_e \mathbf{B}_e \Big) \boldsymbol{\eta}_k = \mathbf{f}_{k+1} - \sum_{e=1}^{M} w_e \mathbf{B}_e^T \boldsymbol{\sigma}_{e,k}^* \tag{10}$$

$$\boldsymbol{\sigma}_{e,k} = \boldsymbol{\sigma}_{e,k}^* + \mathbb{C}_e \sum_{\alpha=1}^{N} \mathbf{B}_{e\alpha} \boldsymbol{\eta}_{\alpha,k} \tag{11}$$

where $(\boldsymbol{\varepsilon}_{e,k}^*, \boldsymbol{\sigma}_{e,k}^*)$ is the unknown optimal local state in the material data set for material point e.

2.1. Solution algorithm

The simplest Data-Driven solver involves the fixed point iteration:

$$\mathbf{z}_k^{(j+1)} = P_{E_k} P_{D_k} \mathbf{z}_k^{(j)} \tag{12}$$

where j is the iteration number, $P_{D_k}\mathbf{z}_k^{(j)}$ denotes the closest point projection onto D (i.e. finding the point in the material data set that is closest to $\mathbf{z}_k^{(j)}$), and $P_{E_k}\mathbf{y}_k^{(j)}$ denotes the projection of a fixed $\mathbf{y}_k \in D_k$ onto E_k. The algorithm converges when the local states remain unchanged under the closest point projection to the material data set [1], as shown in Algorithm 1.

Algorithm 1 Fixed-point solver

Require: Data sets $D_{e,1}$ (via offline phase space sampling: **Algorithms** 2, 3),
Strain operators $\mathbf{B}_e$ (e=1, ..., M), Applied forces $\mathbf{f}_1$
for all time steps $k = 1, ..., N_k$ **do**
i) Set iteration $j = 0$. Initial local data assignment:
for all $e = 1, ..., M$ **do**
if k = 0 **then**
Choose $(\boldsymbol{\varepsilon}_{e,1}^{*,(0)}, \boldsymbol{\sigma}_{e,1}^{*,(0)})$ randomly from $D_{e,1}$
else
Set $(\boldsymbol{\varepsilon}_{e,k}^{*,(0)}, \boldsymbol{\sigma}_{e,k}^{*,(0)}) = (\boldsymbol{\varepsilon}_{e,k-1}^{*}, \boldsymbol{\sigma}_{e,k-1}^{*})$
end if
end for
ii) Solve Equations (9), (10) for $\mathbf{u}_k^{(j)}$ and $\boldsymbol{\eta}_k^{(j)}$
iii) Compute local mechanical states:
for all $e = 1, ..., M$ **do**
Solve Equations (1), (11) for $\boldsymbol{\varepsilon}_{e,k}^{(j)}$ and $\boldsymbol{\sigma}_{e,k}^{(j)}$
end for
iv) Assign local material states:
for all $e = 1, ..., M$ **do**
Choose $(\boldsymbol{\varepsilon}_{e,k}^{*,(j+1)}, \boldsymbol{\sigma}_{e,k}^{*,(j+1)})$ in $D_{e,k}$ closest to $(\boldsymbol{\varepsilon}_{e,k}^{(j)}, \boldsymbol{\sigma}_{e,k}^{(j)})$
end for
v) Compute global distance d via Eq. (7)
if $d >$ tol **then**
Augment $D_{e,k}$ via on-the-fly sampling: **Algorithm** 4
end if
vi) Test for convergence
if $(\boldsymbol{\varepsilon}_{e,k}^{*,(j+1)}, \boldsymbol{\sigma}_{e,k}^{*,(j+1)}) = (\boldsymbol{\varepsilon}_{e,k}^{*,(j)}, \boldsymbol{\sigma}_{e,k}^{*,(j)})$ **then**
Set $\mathbf{u}_k = \mathbf{u}_k^{(j)}$, $(\boldsymbol{\varepsilon}_{e,k}, \boldsymbol{\sigma}_{e,k}) = (\boldsymbol{\varepsilon}_{e,k}^{(j)}, \boldsymbol{\sigma}_{e,k}^{(j)})$
else
$j \leftarrow j + 1$, **goto** ii)
end if
vi) Compute $D_{e,k+1}$ using Equations (16), (17)
end for

Two key questions arise in a multiscale interpretation of the Data-Driven framework: i) How can we efficiently sample the phase space, offline and on-the-fly, using lower-scale computations?

ii) How we can parametrize the material history given the available data from these computations? We address these questions in the following sections.

3. Phase Space Sampling

The reliability of a multiscale Data-Driven prediction of mechanical behavior is contingent upon the quality of the material data sets compiled from lower-scale computations. The process of generating these data sets is termed phase space sampling. We discuss here two types of methods: 'offline' sampling, which relies on preexisting experimental data or the identification of a target subset of the phase space, and 'on-the-fly' sampling, which does not require any prior information. Regardless of the method of sampling, a model for micromechanical unit-cell calculations is required, which can be furnished for example by Molecular Dynamics (MD), Discrete Element (DEM), or lower-scale Finite Element (FEM) simulation, depending on the material at hand (Fig. 1 b)). The generation of unit cells at a given initial state of a given material has been addressed in multiple studies (e.g. [33] for athermal granular materials or [34] for thermalized MD systems in various ensembles.) The micromechanical model $\mathcal{M}$ takes as input, components of a local state (e.g. $\boldsymbol{\epsilon}$ for a purely strain-driven calculation), and returns the remaining components of the state (e.g. $\boldsymbol{\sigma} = \mathcal{M}(\boldsymbol{\epsilon})$).

3.1. Offline sampling

Within offline sampling, we propose two methods, goal-oriented sampling and minimax sampling, and provide simple algorithms for both. Goal-oriented sampling takes advantage of experimental field data to generate the input to the micromechanical calculations (Fig. 1 c)). In its simplest form, the input is the strain measured (e.g. using DIC) on different parts of a discretized deforming body. Algorithm 2 presents the simple steps involved.

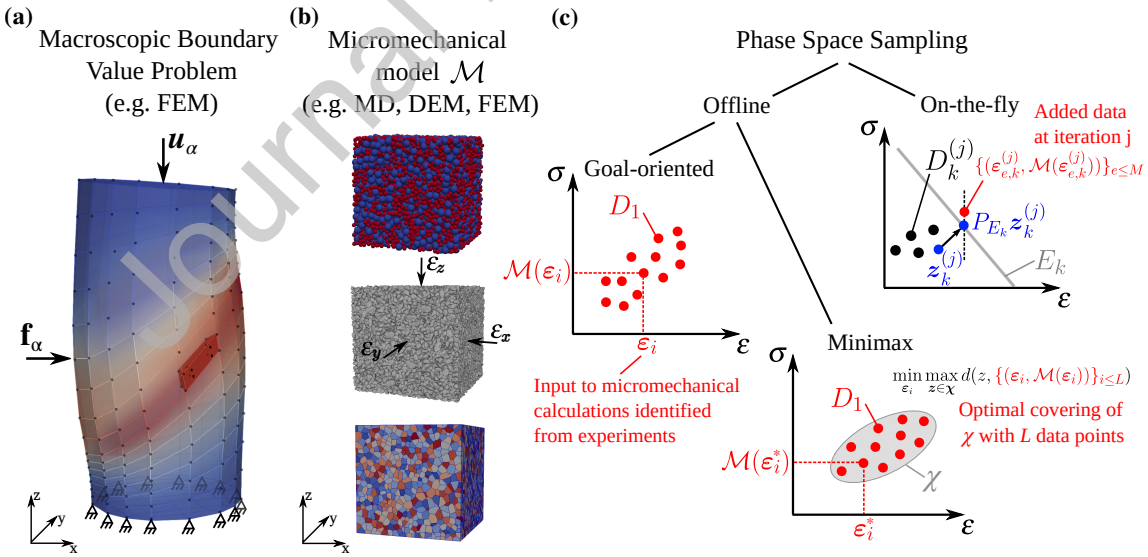

Figure 1: a) Discretized macroscopic boundary value problem b) Micromechanical model for unit-cell calculations during samplings c) Proposed phase space sampling methods within the multiscale Data-Driven framework.

Algorithm 2 Offline goal-oriented phase space sampling

Require: Micromechanical model $\mathcal{M}$, Experimental data
Calculate strains $\boldsymbol{\varepsilon}_i$ on L discrete points of the experimental deformed configuration
for $i = 1, ..., L$ **do**
Carry out unit-cell calculations with imposed $\boldsymbol{\varepsilon}_i$
end for
Set $D_{e,1} = \{(\boldsymbol{\varepsilon}_i, \mathcal{M}(\boldsymbol{\varepsilon}_i))\}_{i \leq L}, e = 1, ..., M$

In the absence of experimental field data, one must resort to alternative sampling techniques. Fortunately, the Data-Driven framework naturally lends itself to optimal minimax sampling [35], due to its formulation in terms of distances in phase space. In this case, we identify a target subset $\boldsymbol{\chi}$ of the phase space (heuristically or using expert knowledge), and seek the optimal input to unit-cell calculations, under computational constraints (number of calculations L), so as to cover $\boldsymbol{\chi}$. Without loss of generality, this input is considered to be the strain (strain-driven calculations), and is determined by the solution of the following problem:

$$\min_{\boldsymbol{\varepsilon}_i} \max_{\mathbf{z} \in \boldsymbol{\chi}} d(\mathbf{z}, \{(\boldsymbol{\varepsilon}_i, \mathcal{M}(\boldsymbol{\varepsilon}_i))\}_{i \leq L}) \tag{13}$$

At the minimizers $\boldsymbol{\varepsilon}_i^*$, we obtain an optimal covering $\{(\boldsymbol{\varepsilon}_i^*, \mathcal{M}(\boldsymbol{\varepsilon}_i^*))\}_{i \leq L}$ of $\boldsymbol{\chi}$, as shown in Fig. 1 c).

Algorithm 3 Offline minimax phase space sampling

Require: Micromechanical model $\mathcal{M}$, Target phase subspace $\boldsymbol{\chi}$
Solve (13) for the minimizers $\{\boldsymbol{\epsilon}_i\}_{i \leq L}$
Set $D_{e,1} = \{(\boldsymbol{\varepsilon}_i, \mathcal{M}(\boldsymbol{\varepsilon}_i))\}_{i \leq L}, e = 1, ..., M$

Remark: This process may involve computationally expensive evaluations of $\mathcal{M}(\varepsilon)$. Therefore, these may be replaced with $\hat{\mathcal{M}}(\varepsilon)$, where $\hat{\mathcal{M}}$ is an appropriate surrogate model or meta-model [36]. We require that the model be simple to evaluate, and able to capture essential aspects of material behavior (e.g. simple plasticity models [37] or machine learning models such as Gaussian Process regression [38] and Neural Networks), and that sufficient preexisting data exist to train the meta-model.

3.2. On-the-fly sampling

It is possible that no experimental data are available, and that we also cannot easily identify the subset of interest in phase space. In this case, we introduce a simple 'on-the-fly' sampling approach, that requires no prior information. The basic principle is the utilization of the Data-Driven solution itself to inform new unit-cell calculations, whenever the distance to the material data set surpasses a chosen threshold. This indicates that the particular region of phase space is not well covered by data and should be targeted by additional calculations. The steps involved are shown in Algorithm 4. As before, without loss of generality, we assume purely strain-driven calculations.

Algorithm 4 On-the-fly phase space sampling

Require: Micromechanical model $\mathcal{M}$, Time step k, Iteration j, Data sets $D_{e,k}^{(j)}$, $e=1,\ldots,M$
  for all $e = 1, ..., M$ **do**
    Carry out unit-cell calculation imposing $\boldsymbol{\varepsilon}_{e,k}^{(j)}$ (obtained in Algorithm 1 - iii))
    Augment local material data set: $D_{e,k}^{(j+1)} \leftarrow D_{e,k}^{(j)} + \{(\boldsymbol{\varepsilon}_{e,k}^{(j)}, \mathcal{M}(\boldsymbol{\varepsilon}_{e,k}^{(j)}))\}$
  end for

Note that, in the applications presented later in this study, and given the presence of experimental data, we shall restrict our attention to offline goal-oriented sampling. The analysis and implementation of the remaining phase space sampling methods will be the subject of a future study.

4. History Parametrization

The efficient and compact history parametrization is a major challenge for materials with an extended memory of their deformation. The first discussion on history parametrization within the Data-Driven formulation is due to Eggersmann et al. [30], where various representational paradigms are explored, including the hereditary/history functional and the internal variable formalism. Alternatively, a thermodynamically-motivated energy based representation, where the local phase space is augmented with the free energy and the dissipation, can be considered. These two parametrizations are particularly appealing within this multiscale paradigm: the energy-based, which is material independent and universal, and the internal variable-based, which must be tailored to a particular material.

4.1. Energy-based parametrization

The energy-based approach effects a parametrization of history by enhancing the state space with the free energy $\mathcal{A}$ and dissipation $\mathcal{D}$, and considering their implicit relation with the remaining state variables $\boldsymbol{\sigma}, \boldsymbol{\epsilon}$. This relation is furnished by the principle of conservation of energy and the second principle (Clausius-Plank inequality):

$$\dot{\mathcal{D}} = \boldsymbol{\sigma} : \dot{\boldsymbol{\epsilon}} - \dot{\mathcal{A}} \geq 0 \tag{14}$$

or, in an time discrete setting, and for material point e:

$$\mathcal{D}_{e,k+1} - \mathcal{D}_{e,k} = \frac{\boldsymbol{\sigma}_{e,k} + \boldsymbol{\sigma}_{e,k+1}}{2} : (\boldsymbol{\varepsilon}_{e,k+1} - \boldsymbol{\varepsilon}_{e,k}) - (\mathcal{A}_{e,k+1} - \mathcal{A}_{e,k}) \geq 0 \tag{15}$$

We can, then, represent the local material data set at time t_{k+1} as:

$$D_{e,k+1} = \{(\boldsymbol{\varepsilon}_{e,k+1}, \boldsymbol{\sigma}_{e,k+1}) \,|\, (\boldsymbol{\varepsilon}_{e,k}, \boldsymbol{\sigma}_{e,k}), (15)\} \tag{16}$$

The above expression implies that the admissible stress-strain pairs at time t_{k+1} (i.e. the pairs looked up by the Data-Driven solver to find the closest point projection to the material data set) are those that are thermodynamically consistent with the previously converged state at time t_k. The special case where $\mathcal{D}_{e,k+1} - \mathcal{D}_{e,k} = 0$ defines a bounded equilibrium set (or elastic domain) on the enhanced phase space. Note that, when dealing with experimental data, the equality above will never be satisfied exactly, which calls for an appropriately defined numerical threshold. As long as the state variables above are easily computable from lower-scale simulations, this parametrization

can be readily obtained.

This is the first data-driven formulation where thermodynamic constraints are explicitly imposed, as opposed to earlier models where these constraints are implicitly imposed, for example using the internal variable formalism [30] or the GENERIC formalism [39].

4.2. Internal-variable based parametrization

The second parametrization under investigation relies on enhancing the state space with a set of internal variables tailored to the material at hand. In this case, the local material data set admits the following representation [30]:

$$D_{e,k+1} = \{(\boldsymbol{\varepsilon}_{e,k+1}, \boldsymbol{\sigma}_{e,k+1}) \,|\, (\boldsymbol{\varepsilon}_{e,k}, \boldsymbol{\sigma}_{e,k}, \boldsymbol{q}_{e,k})\} \tag{17}$$

where $\mathbf{q}_{e,k}$ defines an internal variable - or a collection thereof - encapsulating the material history. In other words, $D_{e,k+1}$ consists of stress-strain pairs in phase space that are attainable from the initial state $(\boldsymbol{\epsilon}_{e,k}, \boldsymbol{\sigma}_{e,k}, \boldsymbol{q}_{e,k})$ or any other sufficiently neighboring state, with respect to an appropriately defined metric. Implicit, here, is the assumption that neighboring states have undergone equivalent history.

Internal variables typically encode information about the microstructure of the material. In several occasions, pertinent internal variables may be known and directly computable from lower-scale simulation data. In the general case, they may be derived through a statistical analysis of the microstructure followed by appropriate dimensionality reduction [40, 41]. The objective is to optimally encode the microstructural information in an internal variable of the lowest possible order, thereby resulting in the lowest-dimensional state space. Most importantly, the Data-Driven paradigm relies only on identifying the relevant variables, while bypassing any need for defininining analytical evolution laws.

5. Application to Granular Materials

Granular materials constitute an excellent candidate for exploring the performance of the proposed framework, since they are known to exhibit complex history-dependent continuum behavior [42]. Traditionally, the quasistatic behavior of granular materials has been described by empirical constitutive laws that are formulated within the framework of plasticity [22, 43–48], and are constrained by a set of principles known as the Critical State Theory [11, 43]. Despite progress in their constitutive description, which was inspired by the recently-enhanced access to grain-scale information (XRCT [49], DEM [7], etc), this conventional modeling approach breaks down during the transition between different regimes, and often relies on prohibitively many parameters [22].

In the remainder of the paper, we will systematically investigate the proposed Data-Driven framework in the modeling of an angular sand. First, we will briefly review the recently developed Level-Set Discrete Element Method which will serve as the machinery for phase space sampling. Then we will explore the performance of the two history parametrizations, within material point simulations involving nonmonotonic loading paths. Finally, we will demonstrate the prediction of granular material behavior against experiments in two boundary value problems.

5.1. In Silico Experiments using LS-DEM

We rely on the recently developed Level-Set Discrete Element Method (LS-DEM) [50] to generate granular material data sets. Similarly to the original formulation of DEM [7], LS-DEM resolves the kinematics of athermal rigid particles interacting through frictional contacts (Fig. 2), but also accounts for accurate particle morphology, described by level set functions[1]. For the purpose of this study, we will assume a linear (Hookean) elastic contact law capped by Coulomb friction, which has been shown to capture all essential aspects of material behavior in sand [52]. Under these assumptions, the interparticle force $\mathbf{f}^c$ associated with a contact c (Fig. 2 c)) is given by:

$$\mathbf{f}^c = \mathbf{f}_n^c + \mathbf{f}_t^c \tag{18}$$

$$\mathbf{f}_n^c = k_n \delta_n \mathbf{n} \tag{19}$$

$$\mathbf{f}_t^c = -\frac{\Delta \mathbf{s}}{\|\Delta \mathbf{s}\|} \min(k_t \|\Delta \mathbf{s}\|, \mu \|f_n\|) \tag{20}$$

where $k_n(k_t)$ is the normal (tangential) Hookean stiffness, related to the elastic properties of the individual particles, δ_n is the interparticle penetration (local contact deformation), $\mathbf{n}$ is the contact normal, $\Delta \mathbf{s}$ the accumulated tangential contact displacement, and μ is the friction coefficient. Note that this contact law (and other more complex laws) may be described in terms of a free energy $\mathcal{A}^c$ and kinetic potential ψ^c (or its associated dissipation function $\mathcal{D}^c$) [33]. This consideration will help us seamlessly describe the energy-based history parametrization. Finally, once all forces acting at a particle $\mathbf{f} = \sum_{c \in \mathcal{C}^p} \mathbf{f}^c$ and resulting moments $\mathbf{m}$ are known, then the particle's dynamics are updated by integrating Newton's equations of motion, given its inertial properties.

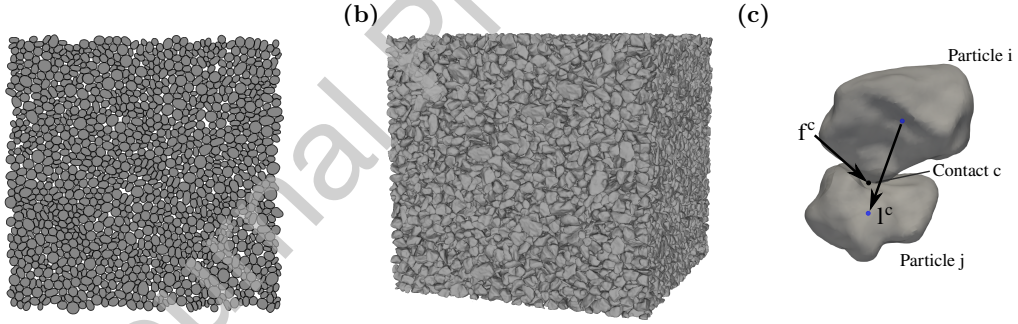

Figure 2: a) 2D and b) 3D granular assembly in LS-DEM. c) Frictional particle contact.

5.2. Exploring History Parametrizations via Material Point Simulations

In this section, we specify the two parametrizations of Section 4 for the case of granular materials, and assess their performance in material point simulations. In the interest of carefully assessing these parametrizations, we will restrict ourselves to previously studied stress paths, instead of using the optimal phase space sampling strategy. Fig. 3 shows the three considered loading-unloading paths, namely isotropic compression, simple shear and constant-volume biaxial compression. For each of these paths, we perform LS-DEM experiments[2] at regularly spaced levels

[1] The level-set approach to DEM dates back to [51]

[2] The term 'experiments' will be used throughout this section to refer to LS-DEM numerical experiments.

of pressure $p = -1/3\ \mathrm{tr}(\sigma)$ or deviatoric stress $q = \sqrt{3/2\boldsymbol{s} : \boldsymbol{s}}$, where $\boldsymbol{s} = \boldsymbol{\sigma} + p\boldsymbol{I}$. At each level, experiments are repeated for multiple samples prepared at the same initial isotropic state. This allows us to generate the data sets that will be used for the subsequent Data-Driven material point predictions. Note that all simulations are non-monotonic, in order to assess the method's ability to distinguish between loading and unloading paths.

We restrict our investigation to quasistatic deformation, in these and all subsequent experiments of this study. This is achieved by ensuring that the dimensionless inertial number $I = \dot{\gamma} d\sqrt{\rho/p}$ is kept below a value of 10^{-4} [53]. Further, in order to avoid boundary effects, and satisfy the Hill-Mandel condition, we enforce periodic boundary conditions [54]. The 2D specimens used for these experiments are comprised of Caicos sand particles characterized during an earlier study [55] with elastic parameters $E = 63.5$ MPa, $\nu = 0.5$ and a friction coefficient of $\mu = 0.5$. An example specimen is shown in Fig. 2 a).

The objective of the Data-Driven material point simulation is to determine the path that minimizes the distance to an applied strain trajectory by navigating through the available data sets in a thermodynamically consistent manner. This translates to a simplification of the Data-Driven problem of Eq. 8 to:

$$\min_{\mathbf{y} \in D_{k+1}} d(\mathbf{z}_{k+1}, \mathbf{y}_{k+1}) \tag{21}$$

where $\mathbf{z}_{k+1}$ is imposed[3], rather than sought from a constraint set, and D_{k+1} is given by Eq. 16 or Eq. 17, depending on the parametrization.

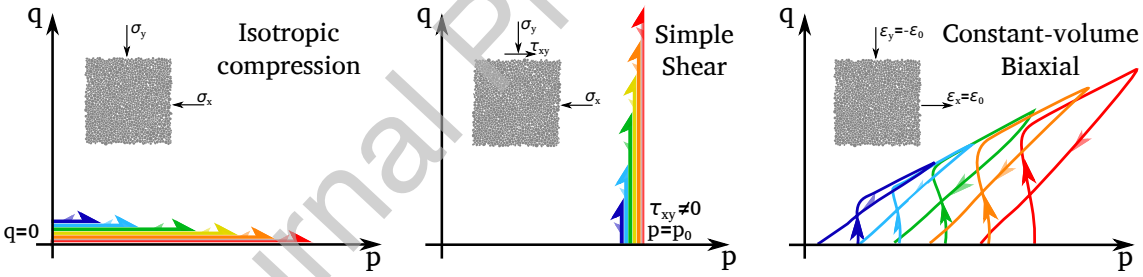

Figure 3: Stress paths for the investigated material point simulations.

5.2.1. Energy-based Parametrization

Adopting the energy-based representation requires access to each of the state variables in Eq. 15. For this granular system, assuming quasistatic conditions, the average stress is given by the Love-Christoffersen expression [56]:

$$\bar{\boldsymbol{\sigma}} = \frac{1}{V} \sum_{c} \mathbf{f}^c \otimes \mathbf{l}^c \tag{22}$$

where the summation runs over each contact c in the assembly. Similarly, the average strain $\bar{\boldsymbol{\epsilon}}$ may be similarly obtained through homogenization [57] or directly from the boundary deformation of the unit cell. Assuming periodic boundary conditions, the Hill-Mandel macrohomogeneity condition [58] is satisfied, allowing us to write:

[3] In fact, only the strain or stress is imposed, depending on whether is is a strain- or stress-driven calculation.

$$d\mathcal{D} = \bar{\boldsymbol{\sigma}} : d\bar{\boldsymbol{\varepsilon}} - d\mathcal{A} \tag{23}$$

where $d\mathcal{D}$ derives from frictional contact dissipation:

$$d\mathcal{D} = \sum_c d\mathcal{D}^c = \frac{1}{V}\sum_c \mathbf{f}_t^c \cdot d\mathbf{u}^{c,slip} \tag{24}$$

where $d\mathbf{u}^{c,slip} = (\mathbf{f}_t^{c,t} - \mathbf{f}_t^{c,t+dt})/k_t$

Finally, the free energy is due to the deformation of contacts under normal and tangential loading:

$$\mathcal{A} = \sum_c \mathcal{A}^c = \frac{1}{2V}\sum_c \left(\frac{\|\mathbf{f_n^c}\|^2}{k_n} + \frac{\|\mathbf{f_t^c}\|^2}{k_t} \right) \tag{25}$$

Let us now explore the performace of this parametrization in material point simulations. Fig. 4 shows the data collected from the isotropic compression experiments up to a pressure of 100 kPa. Different colors represent different sets of simulations, each carried out at a particular value of pressure or deviatoric stress. These represent the data points available to the Data-Driven solver, and are termed 'Data' in the plot legend. The initial compression of the samples is accompanied by a small increase in pressure and a large increase in dissipation, until the jamming transition [59] occurs at $\epsilon_{vol} \approx -0.015$. Compressing beyond that point leads to a significant increase in pressure but only a small increase in dissipation. Similarly, unloading induces negligible change in volumetric strain or dissipation, therefore producing several equilibrium sets (notice the almost horizontal segments in Fig. 4 b)). The same figure shows the sequence of data points determined by the Data-Driven (DD) algorithm for a strain-driven compression cycle at a peak pressure of 50 kPa, that lies inbetween the peak pressures of the available data sets. The success of the data driven simulation is immediately verified upon comparison with a validation experiment at the same peak pressure ('Validation' in the plot legends). Both loading and unloading branches and well captured. Note, in passing, that the difference in terms of dissipation-free energy between different data sets and, hence, between validation and Data-Driven simulation is larger than the same difference in terms of stress-strain. In particular, dissipation is much more sensitive to the initial contact arrangement and leads to larger deviations due to its incremental calculation.

Similarly, Fig. 5 shows the data collected from the simple shear experiments up to 2% shear strain starting from a compressed state of $p_0 = 100$ kPa. Until the samples experience yielding at approximately $\tau = 25$ kPa, the response is largely free of dissipation (Fig. 5 b)) producing once again equilibrium sets. All subsequent loading and unloading branches are tied to significant dissipation. Shown in the same figure, is the Data-Driven simulation carried out to an intermediate peak shear strain $\gamma = 0.01$ and followed by unloading to zero shear stress. Once again, by comparison with a validation experiment based the same shear strain history, it is clear that the algorithm is able to distinguish the best available data set at any point throughout both the loading and unloading branches.

Finally, Fig. 6 show the material data generated by the biaxial compression experiments at various initial isotropic states. The samples initially experience a decrease in pressure followed by a phase transition [22] to a critical state, and, later unload to zero deviator stress. Fig.6 b) shows the applied strain trajectory which is the input to the Data-Driven simulation, and is designed to

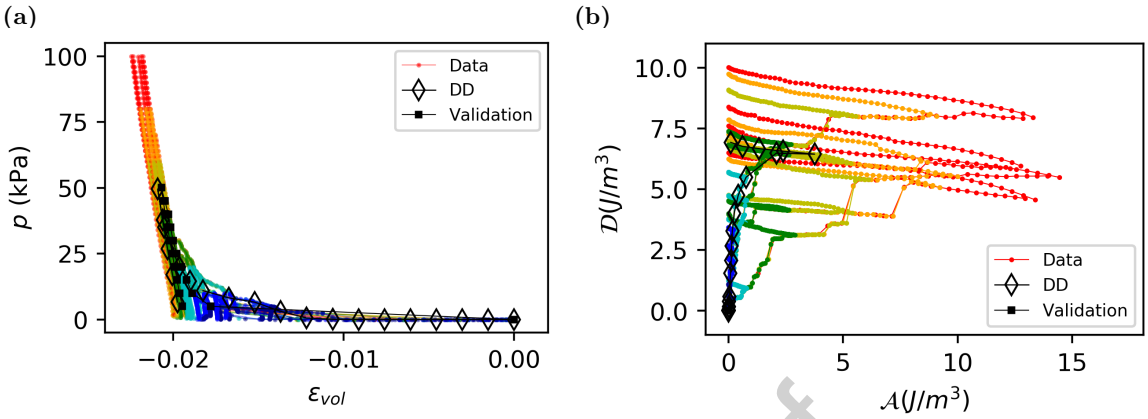

Figure 4: Material data sets (colored), Data-Driven simulation (♢), and validation experiment (■) in isotropic compression: a) Volumetric strain vs Pressure b) Free energy vs Dissipation

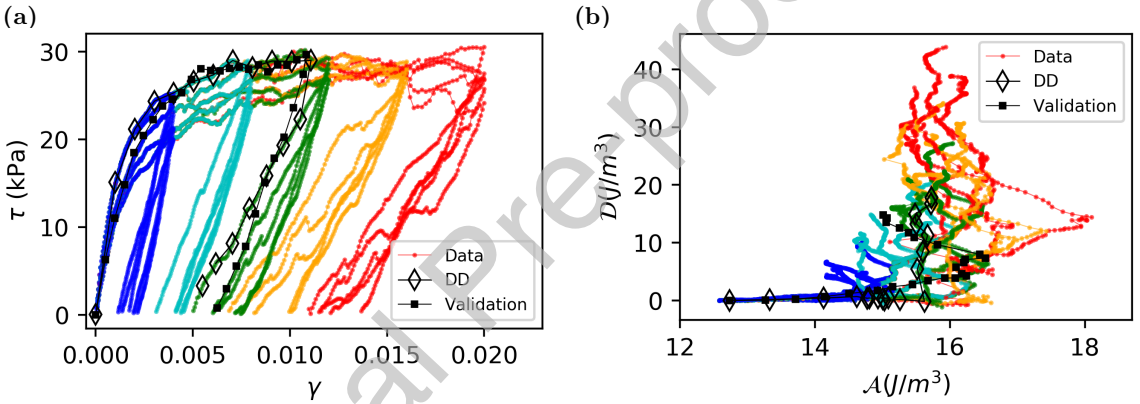

Figure 5: Material data sets (colored), Data-Driven simulation (♢), and validation experiment (■) in simple shear: a) Shear strain vs Shear stress b) Internal energy vs Dissipation

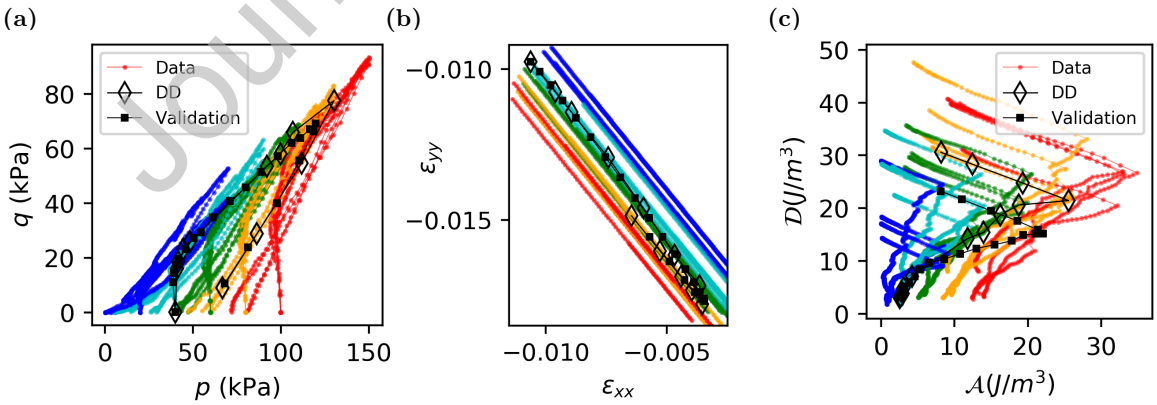

Figure 6: Material data sets (colored), Data-Driven simulation (♢), and validation experiment (■) in constant-volume biaxial compression : a) Pressure vs Stress Deviator b) Strain components ϵ_{xx} vs ϵ_{yy} c) Internal energy vs Dissipation

slightly depart from the constant volume strain trajectories readily available in the data (shown as straight lines). The algorithm is able to transition between sets, to produce a final trajectory in state space that compares well with the validation simulation (Fig. 6 a)).

5.2.2. Internal Variable-based Parametrization

We now turn to the internal variable-based representation of material history. A long-history of studying and modeling granular matter has identified several relevant internal variables such as the packing fraction and fabric tensor [22, 42, 60]. Here, following the discussion in Section 4.2, we derive the internal variables by analyzing the statistics of the internal microstructure, and performing appropriate dimensionality reductions. In this system, the microstructure may be described by the joint probability density of contact normals and forces, $\mathbf{q} = P_{\mathbf{nf}}(\mathbf{n}, \mathbf{f})$ [4]. The former relates to the arrangement of the contact structure, while the combination of normals and forces provides the frictional history. Tractable low-order representations may be achieved through a finite-dimensional approximation of the density. First, restricting our attention to the marginal density $P_{\mathbf{n}}(\mathbf{n})$, and exploiting its symmetry, we can accurately represent it by a second order Fourier expansion [40]:

$$P_{\mathbf{n}}(\mathbf{n}) = \frac{1}{4\pi}(1 + \mathbf{n} \cdot \boldsymbol{\alpha}^n \mathbf{n}) \tag{26}$$

which reduces the problem to the determination of the 6 components of a symmetric contact anisotropy tensor $\boldsymbol{\alpha}^n$. The trace of the tensor $\mathrm{tr}(\boldsymbol{\alpha}^n)$ represents the most important and lowest order description of state which is the packing fraction ϕ, or equivalently the void ratio e [60, 61], followed by the remaining 5 components describing the contact anisotropy [40, 62, 63]. Once $P_{\mathbf{n}}(\mathbf{n})$ is characterized, the description of the internal state is completed by describing the conditional density $P_{\mathbf{f}|\mathbf{n}}(\mathbf{f}|\mathbf{n})$. The simplest way to proceed is through a first-order (mean) description of this density, as $\langle \mathbf{f}(\mathbf{n}) \rangle = \langle f_n(\mathbf{n}) \rangle \mathbf{n} + \langle \mathbf{f}_t(\mathbf{n}) \rangle$, where f_n and $\mathbf{f}_t$ are the normal and tangential force components. For the latter, accurate Fourier expansions are, once again, available [64, 65]:

$$\langle f_n(\mathbf{n}) \rangle = \langle f \rangle (1 + \mathbf{n} \cdot \boldsymbol{\alpha}^{f_n} \mathbf{n}) \qquad \langle \mathbf{f}_t(\mathbf{n}) \rangle = \langle f \rangle (\boldsymbol{\alpha}^{f_t} \mathbf{n} - (\mathbf{n} \cdot \boldsymbol{\alpha}^{f_t} \mathbf{n}) \mathbf{n}) \tag{27}$$

where $\boldsymbol{\alpha}^{f_n}$ and $\boldsymbol{\alpha}^{f_t}$ are the normal and tangential force anisotropy tensors, respectively.
Based on the above, we obtain the following set of relevant internal variables:

$$\mathbf{q} = \{\boldsymbol{\alpha}^n, \boldsymbol{\alpha}^{f_n}, \boldsymbol{\alpha}^{f_t}\} \tag{28}$$

Similarly to the energy-based representation (Section 5.2.1), these state variables $(\boldsymbol{\sigma}, \boldsymbol{\epsilon}, \mathbf{q})$ are directly accessible from the micromechanics. While a notion of hierarchy of these variables was discussed above, a systematic investigation of their relative importance for the macroscopic response is required in order to potentially further reduce the dimensionality of the representation.

Let us now explore the performance of the internal variable-based parametrization in the same material point simulations as before. In each of the three cases (isotropic compression, simple shear, biaxial compression), the corresponding precompiled material data sets are now parametrized by

[4] A more general representation is given by the joint probability density of contact normals, branch vectors and contact forces, $\mathbf{q} = P_{\mathbf{nlf}}(\mathbf{n}, \mathbf{l}, \mathbf{f})$. The combination of contact normals and branch vectors encapsulates additional information about particle shape, and becomes redundant for monodisperse spherical assemblies where $\mathbf{n} \equiv \mathbf{l}$.

the contact fabric-, normal force fabric- and tangential force fabric anisotropy tensors (Eqns. 17,28). The sum of Euclidean (Frobenius) norms of each of the three tensors provides a simple metric for this set of internal variables. Finally, a Data-Driven simulation is carried out and compared to a validation experiment.

Fig. 7 a) compares the Data-Driven simulation of isotropic compression to the corresponding validation experiment. As in the case of the energy-based parametrization, the algorithm is able to predict the pressure response of the isotropically strained sample, during both loading and unloading. Fig. 7 b) shows the evolution of the internal variables in a low-dimensional subspace spanned by the void ratio e and the second invariant of the contact fabric anisotropy tensor α^n. The latter represents the intensity of developing fabric, which remains approximately zero during isotropic compression.

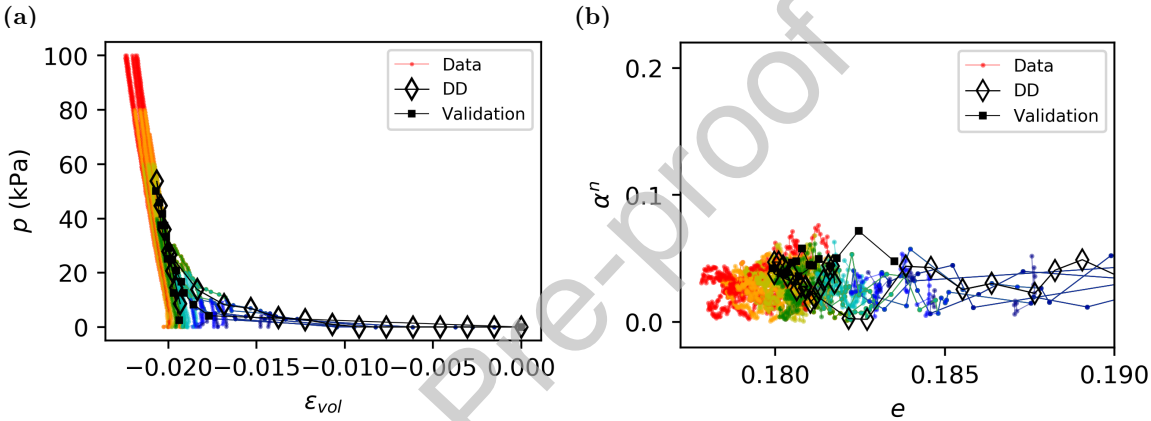

Figure 7: Material data sets (colored), Data-Driven simulation (♢), and validation experiment (■) in isotropic compression: a) Volumetric strain vs Pressure b) Void ratio vs Contact fabric anisotropy invariant

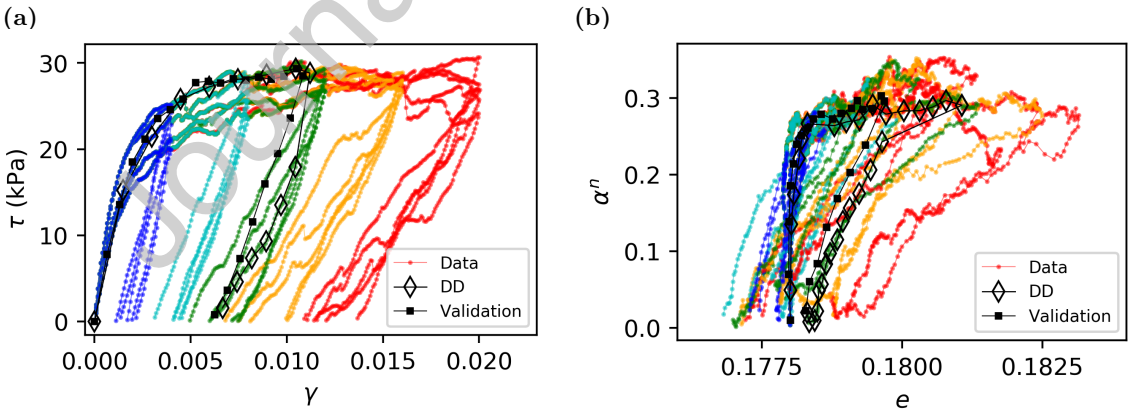

Figure 8: Material data sets (colored), Data-Driven simulation (♢), and validation experiment (■) in simple shear: a) Shear strain vs Shear stress b) Void ratio vs Contact fabric anisotropy invariant

Analogously, Fig. 8 a) compares the Data-Driven simulation of cyclic simple shear to the corresponding validation experiment, exhibiting, once again, a good agreement. Equally satisfactory

agreement is obtained in the internal variable space. Given the size of the space, we choose again to plot only a subspace spanned by e and α_n (Fig. 8 b)). Upon loading, we observe an sharp increase in the contact fabric anisotropy, followed by a smaller increase in the void ratio upon yielding. During unloading, both variables decay to produce an almost isotropic but looser state.

Finally, we assess the performance of the parametrization in the undrained biaxial compression experiments. Fig. 9 a) shows the simulated pressure-stress deviator response to the same strain trajectory as Fig. 6 b), showing excellent agreement with the validation experiment. In terms of internal variables, Fig. 9 b) shows the evolution of the invariants of the contact- and force normal-fabric anisotropy tensors. Despite the departure of the simulated trajectory from the available individual data sets, the algorithm is able to produce consistent transitions between the data sets towards an overall good prediction of the response. This illustrates the interpolating properties of the framework.

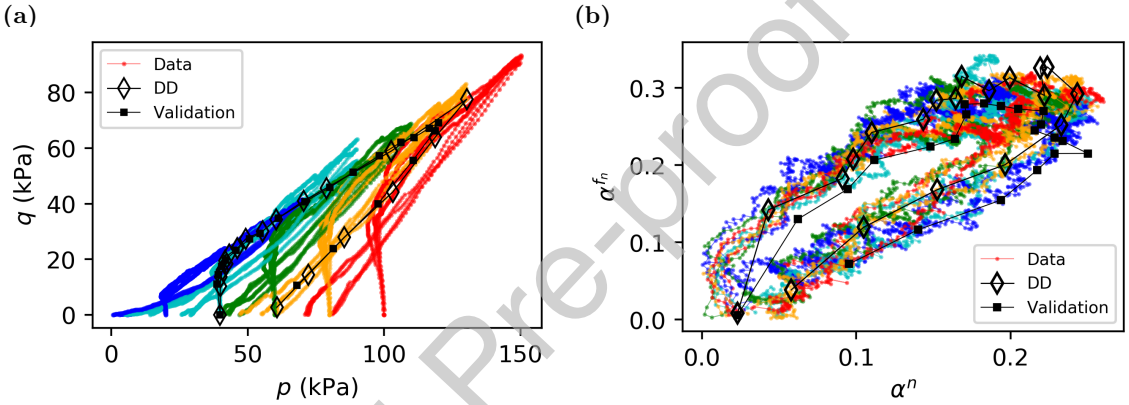

Figure 9: Material data sets (colored), Data-Driven simulation (♢), and validation experiment (■) in biaxial compression: a) Pressure vs Stress Deviator b) Contact fabric anisotropy invariant vs Normal force fabric anisotropy invariant

To summarize, it was shown that both the energy-based and the internal variable-based parametrization perform well in two-dimensional material point simulations of both simple and more complex nonmonotonic stress paths. The first parametrization, furnished by thermodynamics, is universal and material independent. On the other hand, the second parametrization was tailored to granular materials by mathematically describing the microstructure and performing appropriate reductions.

5.3. Boundary Value Problems

We now investigate the performance of the multiscale Data-Driven paradigm, in conjunction with the energy-based history parametrization, in predicting the mechanical response in two boundary value problems: a) plane-strain fault rupture through a dense granular layer, and b) shear banding in three-dimensional triaxial compression. In the first problem, material data sets will be obtained from the experiment itself, in a demonstration of 'self-consistent' phase space sampling. On the other hand, in the second problem, material data sets will be independently generated through unit cell simulations, in a demonstration of the general phase space sampling approach.

330 *5.3.1. Plane Strain Fault Rupture*

We consider a sandbox experiment inspired by the fault rupture study of Anastasopoulos et al. [66]. In the latter, sand is pluviated into a rectangular container to form a specimen of a 10m width and 30m height at prototype scale. A piston underneath the sandbox forces the right part of the specimen to quasistatically subside, inducing a fault rupture through the body of sand at a
335 30° angle to the horizontal (Fig.10 b)).

We employ LS-DEM to simulate the experiment, by replicating the boundary conditions and using the model granular material described in [67]. A snapshot of the LS-DEM simulation along with the resulting contours of dissipation is shown in Fig.10 b). In accordance with similar observations in the physical experiment, both a primary rupture and an antithetic secondary rupture
340 are identified at angles 125° and 52° to the horizontal, respectively.

The virtual specimen is spatiotemporally sampled to produce material data sets - including stress, strain, internal energy and dissipation, in accordance with
345 the energy-based parametrization (Section 4.1). This is achieved by superimposing a finite element mesh on the discrete element assembly, to associate subassemblies to their nearest material point
350 (Gauss integration point). This results in a number of material data sets equal to the number of material points. We refer to this type of phase space sampling as self-consistent since it is in some sense
355 equivalent to the self-consistent Data-Driven identification from field data [68]. In contrast to the goal-oriented sampling discussed in Section 3.1 and implemented in the next application, all components
360 of the local state are readily available, and, hence, no unit cell calculations are carried out. The data sets are initially randomly assigned to each material point in the discretized (FEM) model used for
365 the Data-Driven simulation. Informed by the evolving boundary constraint set, the algorithm is able to make correct associations and transitions between the available data sets, and, thus, capture the mechanical response. Fig.10 c) shows the resulting contours of dissipation in the Data-Driven simulation, which compare well with those of the LS-DEM experiment. In par-
370 ticular, both ruptures are predicted, albeit with angles 122° and 47° respectively. Finally, Fig.11 compares the evolution of surface settlement in the LS-DEM and Data-Driven simulation verifying the good quantitative agreement.

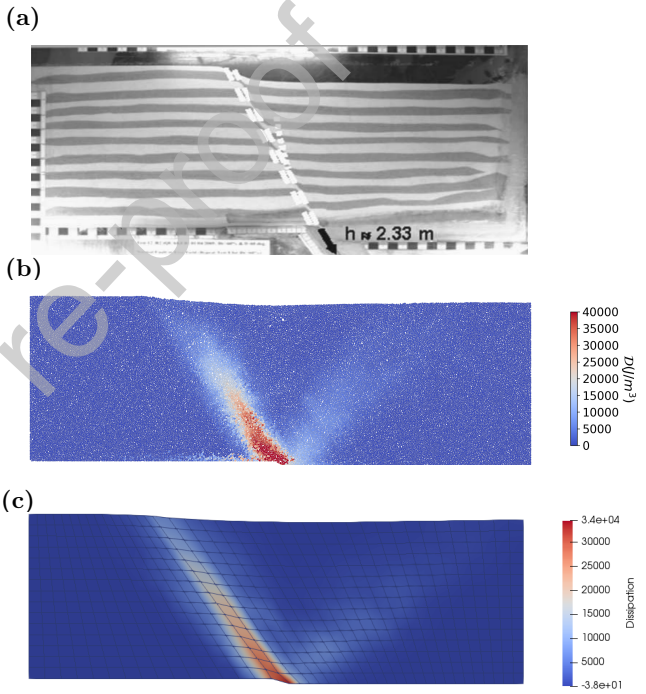

Figure 10: a) Experimental fault rupture setup adopted from Anastasopoulos et al. [66] b) LS-DEM simulation of fault rupture c) Data-Driven simulation of fault rupture and resulting shear bands.

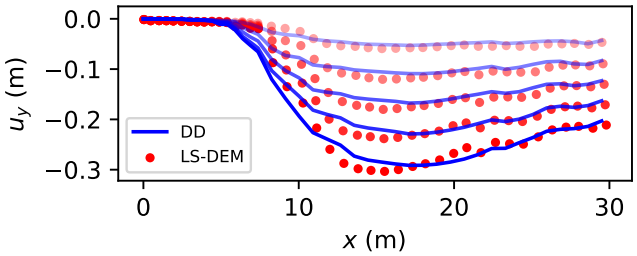

Figure 11: Evolution of surface settlement profile in LS-DEM and Data-Driven simulations.

5.3.2. Shearing Banding in Triaxial Compression

We now consider an *in-situ* triaxial compression experiment on a specimen of angular (Hostun) sand, performed within an XRCT scanner, as described in [69]. Fig. 12 a) shows a 2-dimensional slice of the scanned specimen, which measures 11 mm in diameter and 22 mm in height, and is comprised of 53,939 angular grains. An encasing flexible latex membrane allows the sample to be subjected to radial cell pressure, while a freely rotating platen in contact with the top part of the sample can enforce a vertical compression. The specimen is first compressed isotropically to 100 kPa, and then compressed triaxially by keeping the cell pressure constant while prescribing a vertical displacement to the platen under quasistatic conditions. Eventually, failure occurs through the formation of a persistent shear band.

Recently, the experiment has been modelled using LS-DEM , where, for each physical grain in the sample, an equivalent virtual grain is generated through a level set imaging algorithm [52]. The resulting virtual specimen is, then, subjected to the same boundary conditions, by modeling the membrane as well as the kinematics of the platen. The deformed specimen at critical state is shown in Fig. 12 b). LS-DEM is able to capture both the onset and spatiotemporal evolution of the shear band, as reported in detail in a recent publication [52].

We now simulate the experiment using the Data-Driven framework. In a first simulation, the material data are directly gleaned from the LS-DEM calculation (Fig. 12 b)) in a self-consistent manner, as in Section 5.3.1. In a second simulation, we choose, instead, to compile new material data sets by means of periodic unit cell calculations, using goal-oriented sampling (Section 3.1). To this end, we need to generate cells (Fig. 13 a)) that are representative of the initial local states in the heterogeneous cylindrical specimen, and use them to sample the phase space. First, we compute the void ratio distribution in the specimen, which is enough to fully characterize its heterogeneity given the predominantly isotropic fabric. We then generate cells with void ratios sampled from that distribution (Fig. 13 b)).

Finally, we identify the loading paths to which we shall subject the representative unit cells. We expect that, after triaxial compression to a peak state and the associated onset of shear banding, a bifurcation of the response occurs, with the exterior of the band unloading along the same stress path, and the interior of the band undergoing locally approximately simple shear. Indeed, we can identify these paths via a simple clustering analysis of the experimental data. In particular, we employ the density-based spatial clustering algorithm DB-SCAN [70] equipped with the following distance:

$$d(\mathbf{z}, \mathbf{y}) = |\{\mathbf{z}_k - \mathbf{y}_k\}_{k=1}^{N}| \tag{29}$$

arising from the history-matching metric:

$$|\{\mathbf{z}_k\}_{k=1}^{N}| = \Big(\sum_{k=1}^{n} |\mathbf{z}_k|^2\Big)^{1/2} \tag{30}$$

where $|\mathbf{z}_n|$ is given by Eq. 5). The above metric induces trajectory clustering on the data, which are plotted in terms of isotropic and deviatoric invariants in Fig. 13 c). Two major clusters are obtained, one inside the shear band (red), characterized by pronounced dissipation, and one outside the shear band (blue), where the response remains almost 'elastic'. Sampling 10 paths within each cluster for a total of 20 unit cell calculations is enough to produce the required data sets. In case we encounter a solution that lies unacceptably far from the data set at any point, then additional sampling can be carried out.

In both types of simulations discussed above, the material data sets are parametrized following the energy-based approach, while FEM is used for the discretization of the macroscopic boundary value problem. Note that, in this application, localization arises from the heterogeneity of the

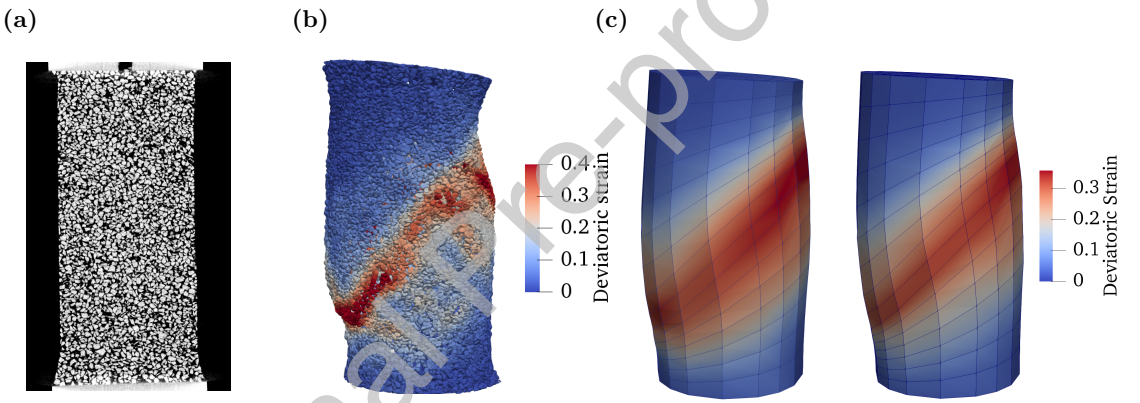

Figure 12: a) XRCT of Hostun sand specimen at the end of isotropic compression. b) LS-DEM simulation, and c) Data-Driven simulation of triaxial compression (left: goal-oriented, right: self-consistent), both at critical state

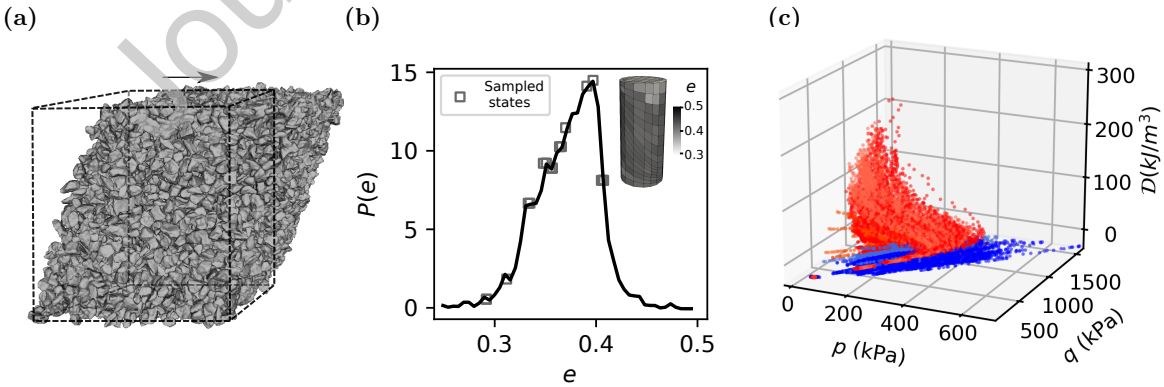

Figure 13: a) Example periodic cell used for phase space sampling b) Initial heterogeneous local void ratio distribution in the cylindrical specimen c) Data clustering in $p-q-\mathcal{D}$ space

sample, rather than by a prescribed discontinuous boundary displacement (as in the fault rupture experiment). Hence, we need to provide partial information about the heterogeneous structure, in order to aid the convergence of the algorithm, and the onset of localization. This is achieved by attaching to a given material point, an initial material data set that corresponds to a void ratio similar to the initial void ratio of that material point in the experiment. This simply serves as a starting point for the fixed point iterations of the solver. Upon this initialization, all material data sets are available to each material point subject to thermodynamic constraints.

Fig. 12 c) shows the deformed specimen obtained from the self-consistent as well as the goal-oriented Data-Driven simulation at critical state. Upon comparison with the deformed LS-DEM specimen (Fig. 12 b)), it is clear that the localized response is well captured. Finally, Fig. 14 compares the macroscopic response in terms of axial strain, principal stress ratio and volumetric strain of the sample, as obtained from the experiment, LS-DEM and the two Data-Driven simulations, all showing good agreement.

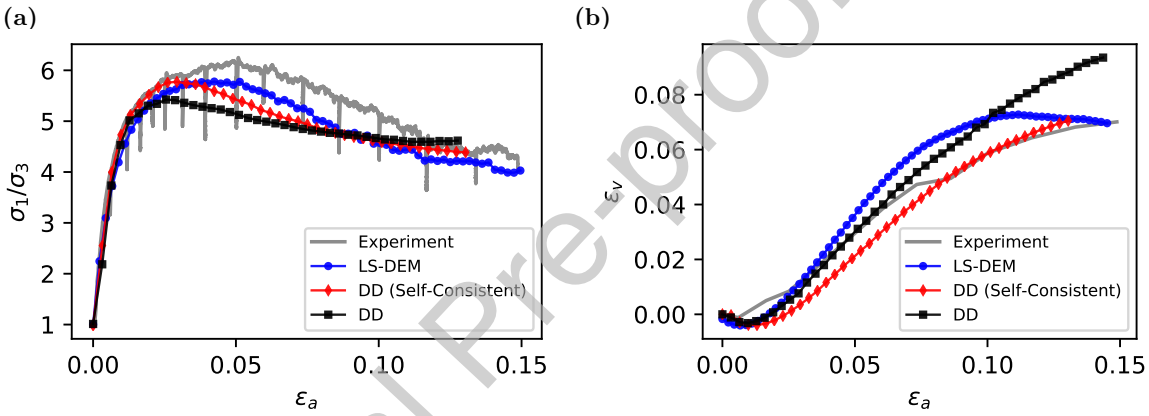

Figure 14: a) Comparison between physical experiment, LS-DEM simulation and Data-Driven prediction (self-consistent and goal-oriented) in terms of a) axial strain vs principal stress ratio and b) axial strain vs volumetric strain

6. Conclusions

We have presented a new multiscale paradigm for the prediction of material behavior. The proposed framework relies on the general inelastic formulation of Data-Driven Mechanics, where the sampling of the phase space is achieved through lower-scale computations, carried out offline or on-the-fly. The resulting material data-sets are appropriately parametrized to account for history-dependence. We have demonstrated an application of the framework to granular materials, where grain-scale computations with the Level-Set Discrete Element Method were used to generate the required data sets. Within this application, we investigated the performance of two history parametrizations: one thermodynamics-motivated and material-independent, and one internal variable-based and tailored to granular materials. The framework was able to capture the nonmonotonic and localized response in both plane-strain rupture and triaxial compression boundary value problems.

It is worth briefly contrasting Data-Driven computing to machine learning-based methods and conventional multiscale methods. Machine learning almost always requires specifying ad hoc effective coordinates, or 'features' , which carry most of the information in the data. The identification of these features is also often ad hoc, and reduces the overall modeling process to old-fashioned fitting or regression. In addition, the internal representation of the machine-learned model is often hidden and devoid of physical meaning. By contrast, Data-Driven methods use the data, all the data and nothing but the data in making predictions. In particular, no empirical or subscale information is lost or manipulated in any way. The physical interpretation of the Data-Driven solutions (that the admissible equilibrium and compatible states should be as close to the data as possible) makes eminent physical sense, and the process of solution is transparent, with no hidden representations or manipulations. The proposed Data-Driven multiscale paradigm are also superior to conventional multiscale methods since the data are readily reusable for future predictions involving the same material. Since searching a material database is typically much faster than a new unit-cell calculation, as more and more data relevant to a particular application are present in the dataset, the framework becomes increasingly superior to conventional techniques.

To conclude, Data-Driven techniques will likely continue to gain popularity in an era when data from high-fidelity simulations and high-resolution experiments are becoming increasingly abundant, as long as these methods serve the long-standing objective of minimizing uncertainty and improving predictive capability. However, techniques such as the proposed framework should not be viewed as a replacement for classical constitutive modeling paradigms that have matured over the years, but rather as a complement in an emerging era of Data-Driven computing. Uncovering and mathematically describing the laws that emerge from the collective behavior of complex micromechanical systems remains a fundamental research goal, and Data-Driven techniques can only help towards that goal, by revealing how much of the rich micromechanical data is relevant to the macroscopic behavior.

7. Acknowledgements

Partial support for this research was provided by US ARO funding through the Multidisciplinary University Research Initiative (MURI) Grant No. W911NF-19-1-0245. This support is gratefully acknowledged.

References

References

[1] T. Kirchdoerfer, M. Ortiz, Data-driven computational mechanics, Computer Methods in Applied Mechanics and Engineering 304 (2016) 81 – 101.

[2] B. D. Coleman, M. E. Gurtin, Thermodynamics with internal state variables, The Journal of Chemical Physics 47 (1967) 597–613.

[3] J. Rice, Continuum mechanics and thermodynamics of plasticity in relation to microscale deformation mechanisms, Division of Engineering, Brown University, 1974.

[4] J. Lubliner, Plasticity theory, Dover, 1990.

[5] F. F. Abraham, D. Schneider, B. Land, D. Lifka, J. Skovira, J. Gerner, M. Rosenkrantz, Instability dynamics in three-dimensional fracture: An atomistic simulation, Journal of the Mechanics and Physics of Solids 45 (1997) 1461 – 1471.

[6] V. Bulatov, F. Abraham, L. Kubin, S. Yip, Connecting atomistic and mesoscale simulations of crystal plasticity, Nature 391 (1998) 669 – 672.

[7] P. A. Cundall, O. D. L. Strack, A discrete numerical model for granular assemblies, Géotechnique 29 (1979) 47–65.

[8] C. Zhou, S. B. Biner, R. LeSar, Discrete dislocation dynamics simulations of plasticity at small scales, Acta Materialia 58 (2010) 1565 – 1577.

[9] R. Madec, B. Devincre, L. P. Kubin, From dislocation junctions to forest hardening, Phys. Rev. Lett. 89 (2002) 255508.

[10] D. E. Spearot, K. I. Jacob, D. L. McDowell, Non-local separation constitutive laws for interfaces and their relation to nanoscale simulations, Mechanics of Materials 36 (2004) 825 – 847.

[11] X. S. Li, Y. F. Dafalias, Anisotropic critical state theory: Role of fabric, Journal of Engineering Mechanics 138 (2012) 263–275.

[12] T. I. Zohdi, J. Oden, G. J. Rodin, Hierarchical modeling of heterogeneous bodies, Computer Methods in Applied Mechanics and Engineering 138 (1996) 273 – 298.

[13] F. Feyel, J.-L. Chaboche, Fe2 multiscale approach for modelling the elastoviscoplastic behaviour of long fibre sic/ti composite materials, Computer Methods in Applied Mechanics and Engineering 183 (2000) 309 – 330.

[14] J. E. Andrade, X. Tu, Multiscale framework for behavior prediction in granular media, Mechanics of Materials 41 (2009) 652 – 669. Advances in the Dynamics of Granular Materials.

[15] M. Nitka, G. Combe, C. Dascalu, J. Desrues, Two-scale modeling of granular materials: a dem-fem approach, Granular Matter 13 (2011) 277–281.

[16] S. Kohlhoff, P. Gumbsch, H. F. Fischmeister, Crack propagation in b.c.c. crystals studied with a combined finite-element and atomistic model, Philosophical Magazine A 64 (1991) 851–878.

[17] E. B. Tadmor, M. Ortiz, R. Phillips, Quasicontinuum analysis of defects in solids, Philosophical Magazine A 73 (1996) 1529–1563.

[18] J. Q. Broughton, F. F. Abraham, N. Bernstein, E. Kaxiras, Concurrent coupling of length scales: Methodology and application, Phys. Rev. B 60 (1999) 2391–2403.

[19] G. J. Wagner, W. K. Liu, Coupling of atomistic and continuum simulations using a bridging scale decomposition, Journal of Computational Physics 190 (2003) 249 – 274.

[20] R. A. Regueiro, B. Yan, Concurrent multiscale computational modeling for dense dry granular materials interfacing deformable solid bodies, in: R. Wan, M. Alsaleh, J. Labuz (Eds.), Bifurcations, Instabilities and Degradations in Geomaterials, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 251–273. doi:10.1007/978-3-642-18284-6_14.

[21] F. Roters, P. Eisenlohr, L. Hantcherli, D. Tjahjanto, T. Bieler, D. Raabe, Overview of constitutive laws, kinematics, homogenization and multiscale methods in crystal plasticity finite-element modeling: Theory, experiments, applications, Acta Materialia 58 (2010) 1152 – 1211.

[22] Y. F. Dafalias, M. T. Manzari, Simple plasticity sand model accounting for fabric change effects, Journal of Engineering Mechanics 130 (2004) 622–634.

[23] J. Ghaboussi, J. H. Garrett, X. Wu, Knowledge-based modeling of material behavior with neural networks, Journal of Engineering Mechanics 117 (1991) 132–153.

[24] R. E. Jones, J. A. Templeton, C. M. Sanders, J. T. Ostien, Machine learning models of plastic flow based on representation theory, Computer Modeling in Engineering & Sciences 117 (2018) 309–342.

[25] N. Vlassis, R. Ma, W. Sun, Geometric deep learning for computational mechanics part i: Anisotropic hyperelasticity, 2020.

[26] K. Wang, W. Sun, A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning, Computer Methods in Applied Mechanics and Engineering 334 (2018) 337 – 380.

[27] M. Mozaffar, R. Bostanabad, W. Chen, K. Ehmann, J. Cao, M. A. Bessa, Deep learning predicts path-dependent plasticity, Proceedings of the National Academy of Sciences 116 (2019) 26414–26420.

[28] T. Kirchdoerfer, M. Ortiz, Data driven computing with noisy material data sets, Computer Methods in Applied Mechanics and Engineering 326 (2017) 622 – 641.

[29] T. Kirchdoerfer, M. Ortiz, Data-driven computing in dynamics, International Journal for Numerical Methods in Engineering 113 (2018) 1697–1710.

[30] R. Eggersmann, T. Kirchdoerfer, S. Reese, L. Stainier, M. Ortiz, Model-free data-driven inelasticity, Computer Methods in Applied Mechanics and Engineering 350 (2019) 81 – 99.

[31] R. Ibaez, E. Abisset-Chavanne, J. V. Aguado, D. Gonzalez, E. Cueto, F. Chinesta, A manifold learning approach to data-driven computational elasticity and inelasticity, Archives of Computational Methods in Engineering 25 (2018) 47–57.

[32] Q. He, J.-S. Chen, A physics-constrained data-driven approach based on locally convex reconstruction for noisy database, Computer Methods in Applied Mechanics and Engineering 363 (2020) 112791.

[33] K. Karapiperis, J. Harmon, E. And, G. Viggiani, J. E. Andrade, Investigating the incremental behavior of granular materials with the level-set discrete element method, Journal of the Mechanics and Physics of Solids 144 (2020) 104103.
[34] G. Fiorin, M. L. Klein, J. Hnin, Using collective variables to drive molecular dynamics simulations, Molecular Physics 111 (2013) 3345–3362.
[35] M. Johnson, L. Moore, D. Ylvisaker, Minimax and maximin distance designs, Journal of Statistical Planning and Inference 26 (1990) 131 – 148.
[36] T. Simpson, J. Poplinski, J. Allen, Metamodels for computer-based engineering design: Survey and recommendations, Engineering with Computers 17 (2001) 129–150.
[37] D. C. Drucker, W. Prager, Soil mechanics and plastic aanalysis or limit design, Quarterly of Applied Mathematics 10 (1952) 157–165.
[38] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, Statist. Sci. 4 (1989) 409–423.
[39] D. González, F. Chinesta, E. Cueto, Thermodynamically consistent data-driven computational mechanics, Continuum Mechanics and Thermodynamics 31 (2019) 239–253.
[40] L. Rothenburg, R. J. Bathurst, Analytical study of induced anisotropy in idealized granular materials, Géotechnique 39 (1989) 601–614.
[41] W. M. Brown, S. Martin, S. N. Pollock, E. A. Coutsias, J.-P. Watson, Algorithmic dimensionality reduction for molecular structure analysis., J Chem Phys. 129 (2008) 064118.
[42] F. Radjai, J. Roux, A. Daouadji, Modeling granular materials: Century-long research across scales, Journal of Engineering Mechanics 143 (2017) 04017002.
[43] K. H. Roscoe, A. N. Schofield, C. P. Wroth, On the yielding of soils, Géotechnique 8 (1958) 22–53.
[44] M. Schofield, C. Wroth, Critical State of Soil Mechanics, McGraw-Hill: London, 1968.
[45] K. Been, M. G. Jefferies, A state parameter for sands, Géotechnique 35 (1985) 99–112.
[46] H. B. Mühlhaus, I. Vardoulakis, The thickness of shear bands in granular materials, Géotechnique 37 (1987) 271–283.
[47] M. Ortiz, A. Pandolfi, A variational cam-clay theory of plasticity, Computer Methods in Applied Mechanics and Engineering 193 (2004) 2645 – 2666. Computational Failure Mechanics for Geomaterials.
[48] J. E. Andrade, R. I. Borja, Capturing strain localization in dense sands with random density, International Journal for Numerical Methods in Engineering 67 (2006) 1531–1564.
[49] J. Desrues, R. Chambon, M. Mokni, F. Mazerolle, Void ratio evolution inside shear bands in triaxial sand specimens studied by computed tomography, Géotechnique 46 (1996) 529–546.
[50] R. Kawamoto, E. Andò, G. Viggiani, J. E. Andrade, Level set discrete element method for three-dimensional computations with triaxial case study, Journal of the Mechanics and Physics of Solids 91 (2016) 1 – 13.
[51] G. Houlsby, Potential particles: a method for modelling non-circular particles in dem, Computers and Geotechnics 36 (2009) 953 – 959.
[52] R. Kawamoto, E. Andò, G. Viggiani, J. E. Andrade, All you need is shape: Predicting shear banding in sand with LS-DEM, Journal of the Mechanics and Physics of Solids 111 (2018) 375 – 392.
[53] G. MiDi, On dense granular flows, The European Physical Journal E 14 (2004) 341–365.
[54] C. O'Sullivan, Particulate discrete element modelling: a geomechanics perspective, CRC Press: New York, 2011.
[55] E. Marteau, J. E. Andrade, A novel experimental device for investigating the multiscale behavior of granular materials under shear, Granular Matter 19 (2017) 77.
[56] J. Christoffersen, M. M. Mehrabadi, S. Nemat-Nasser, A micromechanical description of granular material behavior, Journal of Applied Mechanics 48 (1981) 339–344.
[57] K. Bagi, Analysis of microstructural strain tensors for granular assemblies, International Journal of Solids and Structures 43 (2006) 3166 – 3184.
[58] R. Hill, The essential structure of constitutive laws for metal composites and polycrystals, Journal of the Mechanics and Physics of Solids 15 (1967) 79 – 95.
[59] T. Majmudar, M. Sperl, R. Behringer, Jamming transition in granular systems, Phys. Rev. Lett. 98 (2007).
[60] F. Radjai, J.-Y. Delenne, E. Azéma, S. Roux, Fabric evolution and accessible geometrical states in granular materials, Granular Matter 14 (2012) 259–264.
[61] M. Madadi, O. Tsoungui, M. Ltzel, S. Luding, On the fabric tensor of polydisperse granular materials in 2d, International Journal of Solids and Structures 41 (2004) 2563 – 2580.
[62] M. Satake, Fabric tensor in granular materials, in: P. A. Vermeer, H. J. Luger (Eds.), Proceedings of the IUTAM symposium on deformation and failure of granular materials, A.A. Balkem, Amsterdam, 1982, pp. 63–68.
[63] K.-I. Kanatani, Distribution of directional data and fabric tensors, International Journal of Engineering Science

22 (1984) 149 – 164.

[64] R. J. Bathurst, L. Rothenburg, Observations on stress-force-fabric relationships in idealized granular materials, Mechanics of Materials 9 (1990) 65 – 80.

[65] I. Srivastava, L. E. Silbert, G. S. Grest, J. B. Lechman, Flow-arrest transitions in frictional granular matter, Phys. Rev. Lett. 122 (2019) 048003.

[66] I. Anastasopoulos, G. Gazetas, M. F. Bransby, M. C. R. Davies, A. E. Nahas, Fault rupture propagation through sand: Finite-element analysis and validation through centrifuge experiments, Journal of Geotechnical and Geoenvironmental Engineering 133 (2007) 943–958.

[67] K. Karapiperis, J. Marshall, J. Andrade, Reduced gravity effects on the strength of granular matter: Dem simulations versus experiments, Journal of Geotechnical and Geoenvironmental Engineering 146 (2020) 06020005.

[68] A. Leygue, M. Coret, J. Rthor, L. Stainier, E. Verron, Data-based derivation of material response, Computer Methods in Applied Mechanics and Engineering 331 (2018) 184 – 196.

[69] E. Andò, S. Hall, G. Viggiani, J. Desrues, P. Bésuelle, Grain-scale experimental investigation of localised deformation in sand: a discrete particle tracking approach, Acta Geotechnica 7 (2012) 1–13.

[70] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD96, AAAI Press, 1996, p. 226231.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Konstantinos Karapiperis: Conceptualization, Methodology, Investigation, Writing. **Laurent Stainier:** Conceptualization, Methodology. **Michael Ortiz:** Conceptualization, Methodology. **José E. Andrade**: Conceptualization, Methodology, Funding acquisition